Advanced Tools for Processing Complex Ontology Web Language (OWL) Data

[1, 2] Hussam H Abuazab
[1]Al Khawarizmi International College, Abu Dhabi & Al Ain, UAE
[2]Electrical and Computer Engineering Department, Université du Québec à Trois-Rivières, Trois-Rivieres, Quebec, Canada
hussamAbuazab@khawarizmi.com; hussam.Abuazab@uqtr.ca

Abstract- **Since the launch of Ontology Web Language on 2004, many papers have been introduced in many layers to process the data stored in Relational Databases and flat data which are understandable by humans, but not by computers. And there is actual success in this direction. Despite this, the need for enriching the OWL class sets is a fundamental factor to guarantee successful processing and representing of data by computers. In this paper new OWL class sets that can be obtained through and manipulated by Boolean operators are introduced. The need for new class sets come in order to process and manage complex OWL data, and to improve the expressive power of OWL language. A particular need for such sets arises in processing complex data involved in ontology for human disease which is being developed. The class sets introduced in this paper include the definition of minusOf, De Morgan's Law, and Auxiliary Identity properties. Also, three derived operations are developed in this paper, which are needed to develop ontology for human disease.**

Keywords- OWL Class Sets DeMorgan's Law; MinusOf; Auxiliary Identity; Implication; Exclusive Or; Equivalence

I. INTRODUCTION

OWL is defined to be compatible with the architecture of the World Wide Web in general, with Semantic Web in particular. On 6th September, 2007, W3C has launched the OWL Working Group. The Working Group produced OWL 2, a W3C Recommendation that refines and extends OWL, the Web Ontology Language [1]. This extension of OWL was the latest development on the class sets definitions in the language. Researchers keep focusing in the methodology of utilizing OWL class sets on building proper ontology. In fact, OWL is rich in constructs of class expression category like *owl:oneOf*, *owl:disjointWith*, *owl:unionOf*, *owl:complementOf*; and in Enumerated values like *owl:dataRange*. Also, Boolean class combinations extended to include *owl:intersectOf*, which plays major role in representing intricacies in ontology [2]. Yet, ontology of diseases [3, 4] gives way to complex data and combinations thereof which is very cumbersome to represent using the existing class definitions and operators, which result in poor representations of ontology. Although there are several reasons for poor ontology described in [5], lack of some class set definition is a reason for this problem. So, the motivation of this paper is to cover the shortfall in the class's definitions available in the language, the shortfall appears as complex data and combinations in many types of ontology. To overcome this shortfall, this paper comes to introduce new classes set. The new classes set will help on minimize the gap between conceptual data schemes and ontology, which rise as hot topic on the march of developing tools for building strong ontology [6]. Also, we will see how the new classes set will help on more advanced mapping between ontology and relational databases. In fact, this paper comes to be the first research on the structure of OWL after the last update on 6 September 2007 by the World Wide Web Consortium (W3C).

In the following section, new class sets are introduced to represent complex data of ontology of diseases. Section III discusses the use of these new sets of classes in diagnosing the presence of symptoms of multiple ailments. Section IV introduces the derived operators. Section V concludes the paper.

II. COMPLEX CLASS SETS

In this section we introduce three new OWL class selectors and their resultant class sets, namely, *minusOf, De Morgan's Law,* and *Auxiliary Identity*.

A. MinusOf Class

The *minusOf* represents the set $\{A - B\} = \forall_x x_i \in A \& x_i \notin B$ where A & B are OWL class sets, for example symptoms of two diseases (Fig. 1).

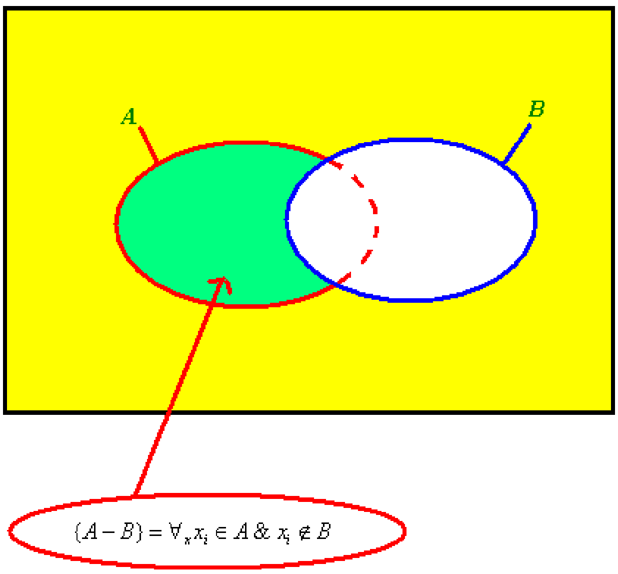

Fig. 1 The class of *minusOf*

A: symptoms of Disease A
B: symptoms of Disease B

So, *minusOf* (A, B) represents the symptoms of disease A excluding those shared with the symptoms of disease B.

B. DeMorgan's Law Class

De Morgan's Law class represents the $\overline{A\cap B\cap C}=\bar{A}\cup\bar{B}\cup\bar{C}$, which implies that the intersect elements are not satisfied. Let us assume that A, B, and C are classes of symptoms of three different diseases, and then $\overline{A\cap B\cap C}$ implies that the common symptoms are not satisfied.

In the way of an example, let us consider the following OWL description of three diseases, for example, acute sinusitis, pneumonia, and common cold. Each disease has known symptoms, and some symptoms may be common among to all or between any two diseases.

```
<owl:Class rdf:ID="Acute_sinusitis">
  <rdfs:subClassOf rdf:resource="#Disease"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#AccompaniedBy"/>
      <owl:someValuesFrom rdf:resource="#Nasal_congestion"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#AccompaniedBy"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:unionOf rdf:parseType="Collection">
            <owl:Class rdf:about="#Bad_Breath"/>
            <owl:Class rdf:about="#swelling_around_eyes"/>
            <owl:Class rdf:about="#swelling_nose"/>
          </owl:unionOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="common_cold">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#AccompaniedBy"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:intersectionOf rdf:parseType="Collection">
            <owl:Class rdf:about="#Bad_Breath"/>
            <owl:Class>
              <owl:unionOf rdf:parseType="Collection">
                <owl:Class rdf:about="#runny_nose"/>
                <owl:Class rdf:about="#Sneezing"/>
              </owl:unionOf>
            </owl:Class>
          </owl:intersectionOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#AccompaniedBy"/>
      <owl:someValuesFrom rdf:resource="#sort_throat"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
```

```
    <owl:Restriction>
      <owl:onProperty rdf:resource="#AccompaniedBy"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:intersectionOf rdf:parseType="Collection">
            <owl:Class>
              <owl:unionOf rdf:parseType="Collection">
                <owl:Class rdf:about="#Cough"/>
                <owl:Class rdf:about="#muffled_voice"/>
                <owl:Class rdf:about="#water_eyes"/>
              </owl:unionOf>
            </owl:Class>

            <owl:Class>
              <owl:unionOf rdf:parseType="Collection">
                <owl:Class rdf:about="#chills"/>
                <owl:Class rdf:about="#weakness"/>
              </owl:unionOf>
            </owl:Class>
          </owl:intersectionOf>
        </owl:Class>
```

After studying the above OWL description, it is clear that there are some common symptoms between the three diseases. Fig. 2 shows the intersections between the three classes of symptoms of the three diseases.

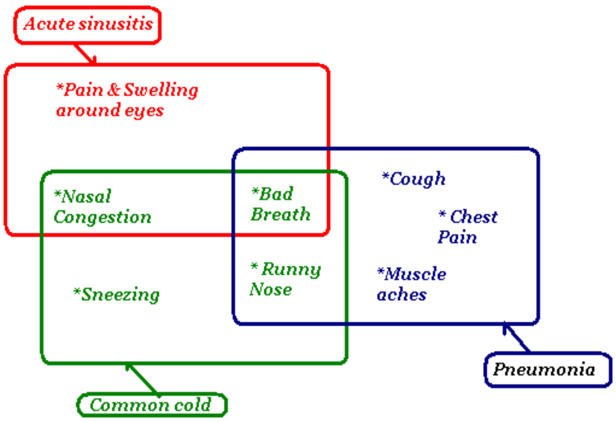

Fig. 2 Class of *De Morgan's Law*

Bad breath is the common symptom among the three diseases. Patient would suffer from bad breath, so he/she may be said to have been infected by at least one or more of the three diseases: acute sinusitis, and pneumonia, and common cold. In case the patient is not suffering from bad breath, this implies that he/she must not have been infected by any of the three diseases.

This can be represented by De Morgan's Law as following:

$$\overline{A\cap B\cap C}=\bar{A}\cup\bar{B}\cup\bar{C}$$

where A is for symptoms of acute sinusitis, B is for pneumonia, and C is for common cold.

So instead of calculating the values of $\bar{A}$, $\bar{B}$, and $\bar{C}$, and then finding the union of these classes, applying De Morgan's Law, it will result in professional solution in means of time and complexity.

C. Auxiliary Classes

Auxiliary Classes represent the sets:

1. $A+AB=A$

By supposing A: symptoms of Disease A, and B: symptoms of Disease B, the set $A+AB=A$ implies that the patient suffers from symptoms of disease A, and the shared symptoms of diseases A and B. So the patient is infected by disease A.

2. $A+\bar{A}B=A+B$

In this case, the patient suffers from symptoms of disease A, and symptoms of disease B, except the ones common with disease A. So the patient is infected by disease A and B.

3. $(A+B)(A+C)=A+BC$

In this case, the patient suffers from symptoms of disease A, and symptoms of disease B. So, let us say, the patient suffers from sneezing, symptom of common cold, and from pain and swelling around eyes, symptom of Acute Sinusitis; also, the patient suffers from chest pain, symptom of Pneumonia, therefore the patient is surely suffering from bad breath which is the intersect symptom of both Acute Sinusitis and Pneumonia (See Fig. 2).

III. DISCUSSION

The mentioned three complex classes comply to the standards of OWL classes regarding the name (URIref) with the "rdf:ID" attribute; also, the class may have properties and instances. The tag for defining the new classes is OWL:Class.

The *minusOf* class is used to implement the cases where the shared symptoms of two diseases are not available while the non-shared symptoms of one disease are available; let us assume a patient is suffering from sneezing, which may imply that he/she is infected by common cold, but the patient is neither suffering from nasal congestion nor bad breath, which implies the patient is not infected by acute sinusitis for sure, as A-B is true for the patient. So the symptoms of pain and swelling around eyes need not be investigated, and the result is diagnosed more precisely and in lesser time. Let us see the reference in Fig. 3.

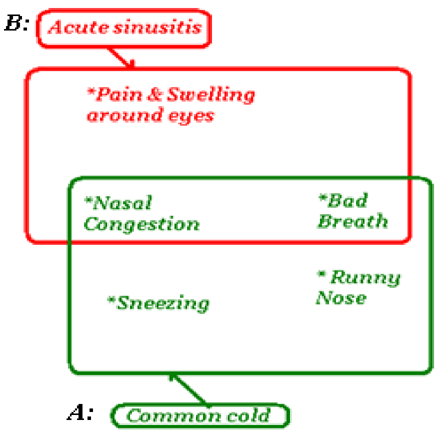

Fig. 3 *MinusOf* Class

The De Morgan's class is used to simplify the diagnosing path, for example, if the patient is not suffering from bad breath then he/she is not infected by either of common cold, acute sinusitis, or pneumonia. This also indicates that one of the other non-common symptoms may be investigated to determine which ailment is present. Let us see the reference in Fig 4.

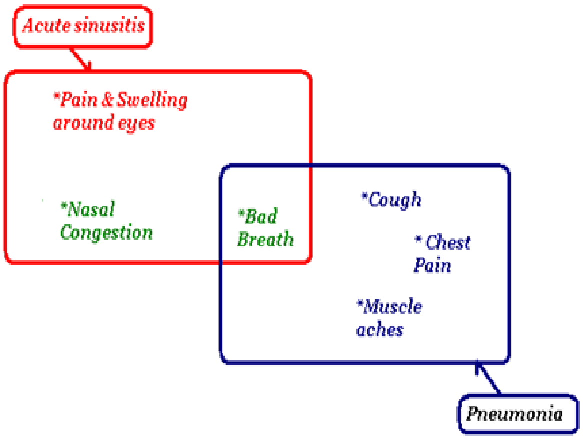

Fig. 4 $\overline{A \cap B} = \bar{A} \cup \bar{B}$

Without using the De Morgan's class, many steps are to take place before proving that the patient who is not suffering from bad breath is not infected by either of common cold, or acute sinusitis, or pneumonia; so if the patient having pneumonia implies that he is suffering from bad breath and from cough and chest pain and muscles aches, which may represented by

A: bad breath,
B: cough,
C: chest pain,
D: muscles ache.
So the patient having pneumonia satisfy the equation:

$$A \cap B \cap C \cap D.$$

While the patient who does not have pneumonia satisfy the equation:

$$\overline{A \cap B \cap C \cap D}.$$

This can be achieved using De Morgan's class, where

$$\overline{A \cap B \cap C \cap D} = \bar{A} \cup \bar{B} \cup \bar{C} \cup \bar{D},$$

We need to prove either one of the symptoms is not applied, so the patient is not suffering from pneumonia.

The *minusOf* Class could have the property of *His Excluded Instance* to represent the instances of class A which are shared with class B. The *DeMorgan'sLaw* Class could have the property *Noncommon Instance* to represent the complement of the common instances between the input classes.

On the *Auxiliary Classes*, and the first class $A+AB=A$, if the patient is suffering from sneezing, symptom of common cold, and from nasal congestion, symptom of both common cold and acute sinusitis, then the patient is infected by common cold for sure, while more investigations are required to decide whether he/she infected by acute sinusitis as well. The second class of the Auxiliary *Classes*, $A+\bar{A}B=A+B$, states that if the patient is suffering from sneezing, symptom of common cold, and from pain and swelling around eyes, symptom of acute sinusitis but not of common cold, then the patient is infected either by common cold or acute sinusitis or both diseases.

The third *Auxiliary Class*, $(A+B)(A+C)=A+BC$, indicates that if the patient is suffering from common cold, disease A, and sometimes suffering from acute sinusitis or pneumonia, so the patient for sure suffering from bad breath.

The web is overloaded with enormous amount of information that is searched based on its syntax [7]. Therefore, poor syntax will result in difficult search and poor searching result. The new class sets will produce strong syntax and more clear ontology. Also, the positive impact of the new class sets is quite clear on mapping the ontology to relational database. Both classes' minusOf and Auxiliary will result in reducing data redundancy, while the class De Morgan's Law will easily find the common attributes between entities and foreign keys, while some studies are still running to evaluate the effect of the new classes on mapping the ontology to relational database.

IV. DERIVED OPERATORS

Three basic operations are defined in OWL, AND, and OR, and NOT. While more operations can be derived from these three basic ones like NAND, NOR, which are already defined in OWL references. In the ontology for human disease, more operations are needed and to be defined. New operations could be derived from Boolean Math, like *implication, exclusive, and equivalence*.

Implication produces a *false* value if and only if the first operand is true and the second operand is false. Implication equation is:

$$p \rightarrow q.$$

And it can be read, if p then q, which is equivalent to:

$$\overline{p} \cup q.$$

Exclusive Or is an operation that produces a *true* value if the both operands are not the same, i.e. one but not both of its operands is true. Exclusive Or equation is:

$$p \text{ xor } q.$$

Equivalence is an operation that produces a *true* value if both operands are the same, i.e. it is inversion of the exclusive Or. Equivalence equation is:

$$p \text{ xnor } q$$

Or

$$p \leftrightarrow q.$$

The *implication* operator can simplify the representation of the ontology. So, if the patient has acute sinusitis, it implies he/she has bad breath. This could be represented as below (Table I):

P: patient gets acute sinusitis;
Q: patient has bad breath.

TABLE I IMPLICATION TRUE TABLE

P	Q	*implication*
T	T	T
T	F	F
F	T	T
F	F	T

So, if the result of implication operator is false, it implies the patient does not get acute sinusitis, as long as he/she has no bad breath.

If the patient gets Pneumonia, it is equivalent to that he/she has chest pain; while if the patient has no chest pain, for sure he/she does not get Pneumonia. For the same definition of the operands P and Q as shown above, the truth table of the equivalence operands is as below (Table II).

TABLE II EQUIVALENCE TRUE TABLE

P	Q	Equivalence
T	T	T
T	F	F
F	T	F
F	F	T

The Exclusive Or is important to develop the fact that the patient who has chest pain must have Pneumonia, so Pneumonia and healthy chest could not come together (Table III).

P: patient gets Pneumonia;
Q: patient has healthy chest.

TABLE III EXCLUSIVE OR TRUTH TABLE

P	Q	Exclusive Or
T	T	F
T	F	T
F	T	T
F	F	F

As shown above, the three new derived operators are needed to build and develop strong, tenacious, and accurate ontology for human disease, which sequences in more proper mapping of ontology to relational database.

V. CONCLUSION

In this paper, new ontology classes are introduced. The *minusOf* class is used for example to represent the symptoms of a disease except those shared with some other disease, which implies that the patient is infected by one but not the other disease. The *DeMorgan's Law* class is utilized to check the availability and non-availability of common symptoms of more than one disease, which implies that the patient is not infected by all diseases causing the shared non-available symptoms. Finally, three auxiliary classes are defined to represent various symptoms' availability.

The *minusOf* and *DeMorgan's Law* classes introduced in this paper will play primary role on easing the representation of complex knowledge, especially medical knowledge, and the knowledge that requires such sort of complex classes. The three auxiliary classes defined are meant to ease the querying process by deducing the common instances and individuals of different classes.New derived operators are defined, *implication*, *exclusive or*, and *equivalence*. These operators played important role to produce well organized ontology.

In the future, more of such specialized classes may be introduced in order to enhance OWL's level to represent more complex knowledge. Also, similarity may be drawn with implementing combination of logic gates that subsume certain

knowledge, and consequently ease the processing of related data sets. Also, more studies need to be conducted to advance the mapping of ontology into relational database to make it more accurate, more flexible, and smoother, either by producing more tools, or by finding out rules.

REFERENCES

[1] (27th October, 2009) W3C Recommendation http://www.w3.org/TR/2009/REC-owl2-overview-20091027/

[2] Bennacer, N. Bourda, Y. Doan, B. "*Formalizing for querying learning objects using OWL*", Supelec, France Sept. 2004; ISBN: 0-7695-2181-9.

[3] Chris Mungall, Georgios Gkoutos, Nicole Washington and Suzanna Lewis" Representing Phenotypes in OWL.

[4] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P.Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M.Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. the gene ontology consortium. Nat Genet, 25(1):25–9, 2000. 1061-4036 Journal Article.

[5] Martin Hepp, Jos de Bruijn GenTax: A Generic Methodology for Deriving OWL and RDF-S Ontologies from Hierarchical Classifications,Thesauri, and Inconsistent Taxonomies Int'l Journal on Semantic Web and Information Systems (IJSWIS), , pp. (2006b).

[6] Jarrar, M., Demey, J., Meersman, R.: On Using Conceptual Data Modeling for Ontology Engineering. In Spaccapietra, S., March, S., Aberer, K., (Eds.): Journal on Data Semantics (Special issue on Best papers from the ER, ODBASE, and COOPIS 2002 Conferences). LNCS, Vol. 2800, Springer, pp.:185-207. October 2003.

[7] Mostafa E. Saleh: Semantic-Based Query in Relational Database Using Ontology: Canadian Journal on Data, Information and Knowledge Engineering Vol. 2, No. 1, January 2011.

Hussam H Abuazab has a PhD in Computer Engineering earned on 2012 from Université du Québec à Trois-Rivières, Trois-Rivieres, Quebec, Canada. He was born in Kingdome of Saudi Arabia, 1972.

He has worked as LECTURER since Sep. 2001 in different universities in United Arab Emirates. He worked on position of senior system analyst for five years in Kingdom of Saudi Arabia. He is currently working on lecturer position in Al Khawarizmi International University College, Abu Dhabi, UAE. He is the coauthor of the book: "Theory and Practice of Cryptography Solutions for Secure Information Systems" (701 E. Chocolate Ave. Hershey, PA 17033, USA, IGI Global, 2012). He published: "Advanced Key Distribution Center (AKDC) Algorithm",(IADIS International Conferences Informatics 2010, Wireless Applications and Computing 2010 and Telecommunications, Networks and Systems 2010 MCCSIS 2010: 26-28 July, 2010 Freiburg Germany) pp 259-262, ISBN: 978-972-8939-19-9 © 2010 IADIS. Also, he published: "An Approach on Database Index Schemas" (The 2003 International Conference on Information and Knowledge Engineering (IKE'03) June 23 - 26, 2003**,** Las Vegas, Nevada, USA. Published on Volume II, ISBN 1-932415-08-4, pp:581-587. He is involved in research area of super computers and parallel processing algorithms and cloud computing.

Dr. Abuazab is a member of Association of Computing Machinery (ACM). He is a reviewer for the 31st Annual International Computer Software and Applications Conference (COMPSAC 2007). He is a member of Internal Technical Committee of Al Khawarizmi International University College, Abu Dhabi & Al Ain Conference in 2012.